Assimilation of one satellite SAR image for flood simulations. Method and test case (Moser river)

Renaud Hostache, Xijun Lai, Jerome Monnier, Christian Puech

Advances in Geosciences
Vol. 22: Atmospheric Science (2010)
Eds. Chun-Chieh Wu *et al.*

ASSIMILATION OF ONE SATELLITE SAR IMAGE FOR FLOOD SIMULATIONS. METHOD AND TEST CASE (MOSEL RIVER)

R. HOSTACHE[†], X. LAI[‡], J. MONNIER[§,*] and C. PUECH[†]
[†]*UMR TETIS, Cemagref Montpellier, France*
[‡]*Toulouse Mathematics Institute & INSA, Toulouse, France*
[§]*Nanjing Institute of Geography & Limnology,*
Chinese Academy of Sciences, Nanjing, P.R. China
[*]*jerome.monnier@insa-toulouse.fr*

In view of improving numerical flood prediction, a variational data assimilation method (4D-var) applied to a 2D shallow water model and using distributed water level obtained from one Synthetic Aperture Radar (SAR) image is presented. The RADARSAT-1 image leads to water levels with a ± 40 cm average vertical uncertainty of a Mosel River flood event (1997, France). Assimilated in the 2D shallow water hydraulic model, these SAR derived spatially distributed water levels prove to be capable of enhancing model calibration. Indeed, the assimilation process can identify some optimal Manning friction coefficients. Moreover, used as a guide for sensitivity analysis, remote sensing water levels allow also in identifying some areas in the floodplain and the channel where Manning friction coefficients are homogeneous. This allows basing the spatial segmentation of roughness coefficient on floodplain hydraulic functioning.

1. Introduction

High spatial resolution SAR spaceborne images allow the estimation of distributed water levels in floodplains with reasonable uncertainty by merging SAR derived flood extent limits with a high-resolution high-accuracy Digital Elevation Model (DEM), see Ref. 1. Furthermore, to be reliable, hydraulic models have to be constrained by using various observed data sets. The model calibration consists in forcing the model to provide outputs as close as possible to observed data by searching optimal values of its parameters. A "hand" calibration of Manning coefficients is often done through trial tests with the use of point observations, such as recorded hydrographs at stream gauges. The variational data assimilation

method based on the optimal control theory of partial differential equations (also called 4D-var method) offers a special powerful tool to fuse in an optimal sense measurements (observations) and the mathematical model. In river hydraulics, variational data assimilation methods have been used successfully for shallow water models.[2–10] The present study aims at investigating whether the variational assimilation of water levels derived from a flood SAR image could help to enhance the calibration of flood inundation models. This would help getting benefits of both variational assimilation and recently developed remote sensing methods, improving flood model calibration.

2. Flood Description and Data Extraction

The area of interest includes a 28 km reach of the Mosel River between Uckange (France) and Perl (Germany). In this area, the Mosel River meanders in a flat plain having an average width of 3 km and a mean slope of 0.05%. It is worth noting the presence of a narrow valley at downflow. The latter behave as a bottleneck during flood events causing upstream water retention area. The propagation velocity of the flood peak in the study area is low, around $2\,\mathrm{km\,h^{-1}}$. The peak discharge recorded at Uckange city stream gauge (upstream boundary) was around $1{,}450\,\mathrm{m^3 s^{-1}}$, corresponding to a 4–5 year time return period.

As hydrometric data, discharge hydrographs were available at three stream gauges located in the study area. These hydrographs are shown in Fig. 1. Recorded discharge hydrographs araise from calculation using observed water stage hydrographs and rating curves: relationship between discharge and water level computed using discharge *in situ* measurements. Considering that higher magnitude discharges have been *in situ* measured in Uckange than in Perl for the rating curve computation, the Uckange hydrograph has been assumed more reliable. As a consequence, only the discharge hydrographs in Uckange and EDF stream gauges are used as ground truth information in this study. Consequently, available ground observations are fairly limited. The time series data of discharges at the EDF gauge station are only available at the beginning and the end of the flood period; the measurements are lacking during the high flood stage because of sensor disability (see Fig. 2).

The SAR image (Systematic Map Image product), amplitude coded, has a pixel spacing of 12.5 m, resulting from the sampling of a complex

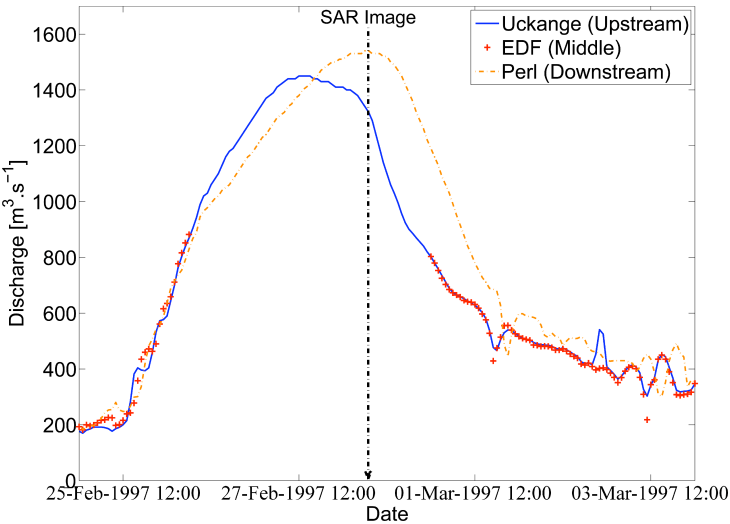

Fig. 1. Recorded hydrographs during the flood event at three different locations.

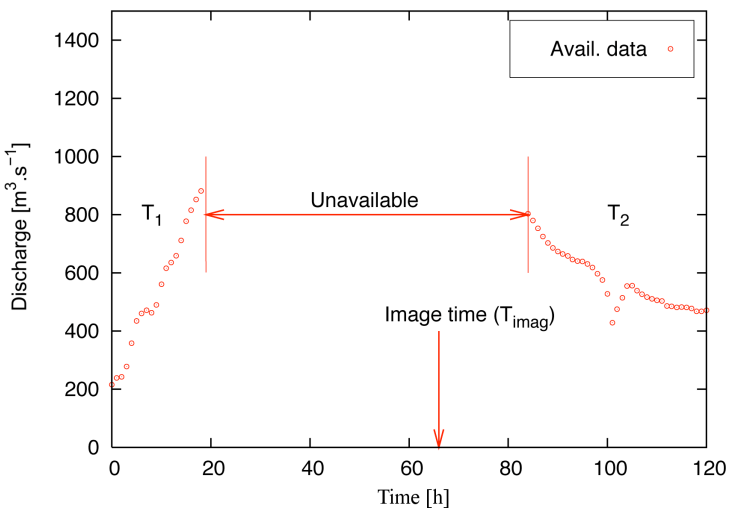

Fig. 2. Observations available: discharge at low water level at EDF gauge station and one (partial) image.

image of 25 m spatial resolution. It was acquired a few hours after the flow peak, at the beginning of the recession, as shown in Fig. 1. In the area of interest, namely the Mosel floodplain, the terrain slopes are rather low and the incidence angle of the RADAR signal does not vary a lot. As a matter

of fact, during the treatment of the SAR image, namely the flood extent delineation, we made use of the pixel's Digital Number (DN).

Furthermore, five air photographs at 1:15,000 scale, flown by the French National Institute of Geography (IGN) in 1999 out of flooding period, have been acquired. They have been used to digitize land use information: buildings, urban areas and sparse habitat, and high vegetation, forests, sparse trees, hedges, etc. Indeed trees and buildings perturb the backscattering signal, so we choiced to mask these areas before processing: the whole process runs using only reliable areas.

The topographic and bathymetric raw data have been provided respectively by the North-Eastern French Navigation Services (SNNE) and the DIREN Lorraine as 3D points and 3D lines — calculated by photogrammetry using air photographs at 1:8,000 scale and SONAR sounding. The average altimetric uncertainty on the raw data is about 25 cm (DIREN information) for the floodplain and 1 cm (SNNE information) for the channel. Using a linear interpolation between points and lines, a Triangular Irregular Network (TIN) Digital Elevation Model (DEM) has been generated and then converted to a RASTER DEM (pixel size: 7 m), easily superimposable with image data. Hereafter, DEM acronym will be assigned to the topographic RASTER data. Moreover steep banks, potentially badly represented, are eliminated from the final process.

3. Water Level Estimation Using SAR Image

Based on the method developed in Ref. 11 providing water level estimates with a ±18 cm mean uncertainty using flood aerial photographs, the water level estimation method used in this study is more detailed in Ref. 12 (see also Ref. 13).

Using the RADARSAT image, this approach provides spatially distributed water level estimates within a ±40 cm mean uncertainty. It can seem strange to reach such accuracy with large pixels (25 m). Nevertheless, more than to pixel size, the accuracy in this process relates to DEM accuracy and the fact that the vertical estimates concern only places with very low slopes and no perturbing items (trees, houses, etc.).

3.1. *Flood extent extraction*

To deal with the radiometric uncertainty, the discrimination of flooded and non-flooded pixels is done using two threshold values. The first threshold

value, $T_{\min}$, aims at detecting only pixels that correspond to water bodies. $T_{\min}$ has been determined as the minimum radiometric value of non-flooded pixels inside grassland areas (outside the floodplain and outside the permanent water surfaces).[12] The second threshold value, $T_{\max}$, aims at detecting all flooded areas, at the risk of detecting in addition non-flooded areas that have a similar radiometric value to the flooded one. $T_{\max}$ has been determined as the maximum radiometric value of water bodies outside the flooded area, using the SAR image pixels located inside the Mirgenbach lake. The thresholding of the SAR image using $T_{\min}$ and $T_{\max}$ provide a flood extent map with fuzzy limits, coded as follows (see Ref. 12), depending on the intensity I of the SAR image pixels: 0, non-flooded ($I > T_{\max}$); 1, flooded ($I < T_{\min}$), 2, fuzzy limit ($\approx$ potentially flooded) ($T_{\min} \leq I \leq T_{\max}$).

The innovative point of the SAR image processing is the analysis of the relevance of the remote sensing-derived flood extent limits for hydraulic purpose, and especially for water level estimation. To estimate water levels, the flood extent limits are merged with the underlying DEM. Using such a merging, any erroneous flood extent limit will lead to errors in water level estimation. Consequently, flood extent limits prone to error have to be identified before the merging. Errors in the flood extent limits are mainly due to emerging objects such as building and high vegetation[12] that may mask water. To treat this potential errors, it has been chosen to remove all SAR derived flood extent limits located in habitat or vegetation areas. Considering that radiometric and spatial uncertainties have been taken into account, the "enlarged" relevant limits are then assumed to include the real flood extent limits (*Hyp. 1*). These are shaped as small patches which are sparsely distributed along the floodplain (Fig. 3).

3.2. *Preliminary water level estimation*

The second part of the process estimates one range of possible water levels $\mathrm{IWLE} = [\mathrm{WL}_{\min}; \mathrm{WL}_{\max}]$ for each relevant patch. To do so, the maximum and the minimum elevation values are first extracted inside each relevant patch using the DEM Z values. Next, the DEM altimetric uncertainty ($\mathrm{uncert}_{\mathrm{DEM}}$) is taken into account by being, respectively, added/subtracted to the maximum/minimum values extracted previously for each relevant patch:

$$\mathrm{IWLE} = [\min(Z_{\mathrm{patch}}) - \mathrm{uncert}_{\mathrm{DEM}}; \max(Z_{\mathrm{patch}}) + \mathrm{uncert}_{\mathrm{DEM}}].$$

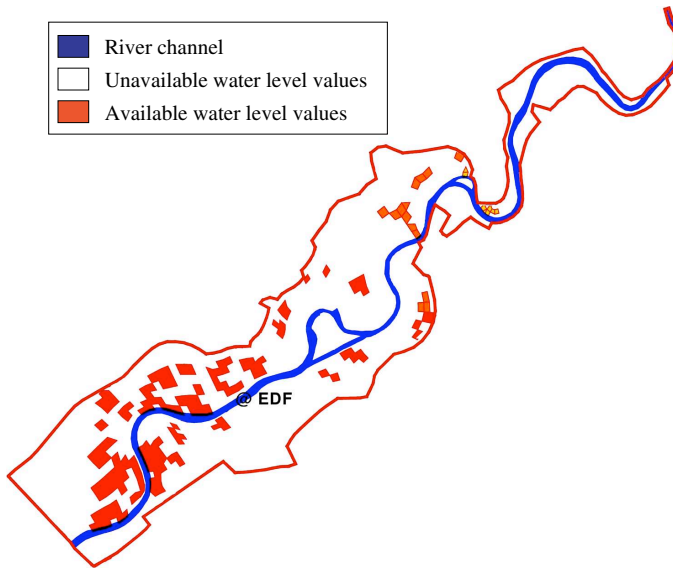

Fig. 3. Spatial distribution of water level values available after image post-treatment.

Since the DEM altimetric uncertainty is taken into account, *Hyp. 1* allows to assume that each range of water level estimation — *IWLE* — includes the real water level (*Hyp. 2*).

3.3. *Final water level estimation*

The last part of the process uses hydraulic rules to constrain the water level estimates and reduce their uncertainty, as proposed firstly by Refs. 11 and 14. In a floodplain, hydraulic laws manage the flow, so that water levels must follow an hydraulic logic: hydraulic energy decreases from upstream to downstream. With low flow velocity, like in the Mosel floodplain, this hydraulic rule can be simplified into a decrease of water level in the flow direction (*Hyp. 3*). To apply *Hyp. 3* on the *IWLE* intervals, flow directions between patches (locations of the water levels) have to be determined. As proposed in Ref. 12, some flow directions between patches have been determined using the shape of the SAR derived flood extent and the lines perpendicular to the elevation contour lines, oriented from the highest to the lowest elevation, called steeper lines hereafter. Furthermore, the main flow directions are assumed to be convergent toward the river channel, following the steeper lines. The steeper lines around relevant patches have

been determined using the contour lines derived from the DEM. Using this criterion, some up-/down-stream relationship between patches have been determined. At the floodplain scale, these relationships constitute an hydraulic hierarchy of the relevant patches. Consequently, according to *Hyp. 3*, the water level must decrease from the patch A to the patch B if A is upstream of B. Due to *Hyp. 2* this induce the following constraints:

$$N_{\max}(B) \leq N_{\max}(A) \text{ (constraint on the maxima)},$$
$$N_{\min}(A) \geq N_{\min}(B) \text{ (constraint on the minima)}.$$

To apply these constraints, the algorithm that has been developed, is flow oriented and impose a decrease on the maxima from upstream to downstream, and vice et versa, an increase on the minima from downstream to upstream. This finally provides intervals of constrained water level estimation IWLE $= [\mathrm{WL}_{\min}; \mathrm{WL}_{\max}]$, with a half mean range of about ± 40 cm.

As a consequence, the method allows the definition of a set of distributed water levels across the floodplain at the satellite overpass time. These remote sensing-derived water levels are then used as new observation for the hydraulic modeling.

As a matter of fact, remote sensing derived data are assumed to be an observation of water depth along the floodplain. The single image is extracted from the satellite image at 6:00 am, Feb 28, 1997. The locations where the water level values have been extracted are shown in Fig. 3.

4. Mathematical Model with Variational Data Assimilation (4D-var)

For more details, we refer to Refs. 13 and 15. Numerical computations are performed by using the DassFlow software.[6]

4.1. *Mathematical model*

The flood flow is modeled by the two-dimensional shallow water equations (2D-SWEs), see Refs. 13 and 15 for more details. The unknowns are h the water depth, $Q_x = hu$ and $Q_y = hv$ the unit discharge in the x- and y-directions ((u, v) is the velocity). The friction slopes are evaluated using the Manning formula.

The full inverse model (with variational data assimilation) includes the forward model (2D SWEs), the adjoint model and the minimization algorithm. The 2D-SWEs is solved using a finite volume method on unstructured meshes. The adjoint model is directly derived from the source codes of forward model by automatic differentiation. The cost function J that has to be minimized is defined in detail below. As a matter of fact, the optimization problem to be solved is: $\min_{\mathbf{p}} J(\mathbf{p})$, where $\mathbf{p}$ is the control variable. Here, the latter is the Manning coefficient (spatially distributed). We refer to Refs. 13 and 15 for more details.

4.2. *Preliminary forward run*

As a preliminary study, the Manning coefficients are set to an *a priori* "reasonable" constant value (equal to 0.025). Then, the simulation results are compared with the available measurements, namely the water stage hydrograph at the middle gauge station (EDF) and the SAR image derived water levels in the floodplain at the satellite passover time.

The computed water level at EDF gauge station is very close to measurements; while in floodplain at image time, the computed water level fits much less with available data. Nevertheless, this preliminary forward run shows that the numerical model can reproduce the flow; and the current solution will be the first guess solution (see next section).

Let us point out that the Manning value imposed (0.025) is the best value we found by hand, i.e., using a classical trial-error approach. Furthermore, we did not improve significantly the result if using the land-use information or not (and imposing classical Manning values related to the land-use considered). We refer to Ref. 13 for more details.

4.3. *Cost function*

The cost function J to be minimized (see Section 4.1) contains three terms: $J_{\text{obs}}(\mathbf{p}) = J_{\text{obs}} + \alpha J_{\text{flux}} + \beta J_{\text{reg}}$ (α and β are weight coefficients to be set).

(1) J_{obs} corresponds to the discrepancy between the observations and the computed flow state.
(2) J_{flux} corresponds to the discrepancy of net mass flux.
(3) J_{reg} is a regularization term for smoothing time-dependent control variables $\mathbf{p}$.

The term $(J_{\text{obs}} + \alpha J_{\text{flux}})$ is defined as follows:

$$J_{\text{obs}} + \alpha J_{\text{flux}} = \frac{1}{2\sigma_z^2} \sum_m \left[z_{T_{\text{imag}}}(x_m, y_m) - z^{\text{obs}}_{T_{\text{imag}}}(x_m, y_m) \right]^2 \left[1 + \alpha \|(u, v)(x_m, y_m)\|^2 \right] + \frac{\gamma_1}{2\sigma_Q^2} \left[\int_0^{T_1} (Q(t) - Q^{\text{obs}}(t))^2 \mathrm{d}t + \int_{T-T_2}^{T} (Q(t) - Q^{\text{obs}}(t))^2 \mathrm{d}t \right]$$

where σ_z and σ_Q are the standard deviations of observations, T_{imag} is the image time, (x_m, y_m) is the position of water level measurements, γ_1 is a weighting coefficient, T_1 and T_2 are the time periods of the discharge measurement available at EDF station (see Fig. 2), T is the assimilation time period.

5. Calibration of Spatially Distributed Manning Friction Coefficients (Land-Cover Based Spatial Distribution)

The Manning friction coefficients n, that represents the resistance to the flow in channels and floodplains, are empirical. In fact we must point out that (n) does not really exists, since it is strongly scale dependent, because it integrates all the friction processes at all scale. This section aims at showing the capability of SAR derived spatially distributed water levels to enhance the Manning coefficient calibration, in comparison to a "hand" calibration using trial-error tests. Thus, the data assimilation method presented here uses "orthogonal data": water heights derived from water limits versus classical water depths in gauge station, see Fig. 2.

In this section, the spatial distribution of Manning friction coefficient is based on land-cover classes as classically done. Various cases have been considered depending on the total number of classes taken into account (between 1 and 10). The domain is decomposed by using land-cover classes, one Manning coefficient value n being set by land-cover class. Four various decompositions have been investigated, depending on the total class number:

(a) A lumped n value: one constant value all over the domain.
(b) Three classes (distributed n) consisting of the main channel and the left and right floodplains.

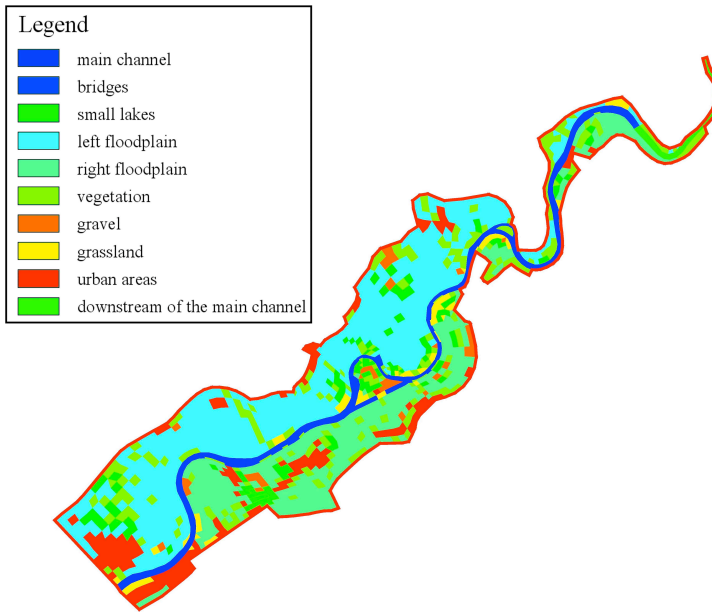

Fig. 4. Land cover: 10 classes (Case d) in the text.

(c) Five classes consisting of the main channel, the bridges (same n value for all bridges), the small lakes, and the left and right floodplains.
(d) Ten classes according to the land-cover classification presented in Fig. 4.

In all numerical experiments, the assimilation period is 66 h from 12:00, Feb 25th (flood event starting time) to 6:00 am, Feb 28, 1997 (SAR image acquisition time).

Identification of Manning coefficients using synthetic data (*twin experiments*). We performed identification experiments based on synthetic data, see Ref. 13 for more details. They showed that the SAR derived spatially distributed water levels and the stream gauge *in situ* measurements can identify properly the spatial distributed Manning coefficients, at least if three land-cover classes only are considered and if the mathematical model is perfect (i.e. observations corresponds to the modeled flow) and no error is introduced in measurements. These last two features are very important. Next, we use the real data and we will notice that the (real) identification problem is much more difficult to solve numerically, even with three land-cover classes only.

Identification of Manning coefficients using real data. In all cases (depending on the land-cover number), the computed value of the main channel is the same whatever the initial guess n_0 (e.g. 0.033, 0.020, 0.025 or 0.050) and the total class number. On the contrary, n in other land-cover classes are almost not modified during the optimization process. This result is consistent with the forthcoming sensitivity analysis (see next section), runs which shows that the total cost function is almost insensitive to the n values outside the main channel.

Using real data, sensitivities can become different, and the identification problems become much more difficult. In the present case, the few sensitivity of the model to the Manning coefficient outside the channel is mainly due to the rather low time return period of the investigated flood event (4–5 years). The results show that the water levels in the floodplain is more influenced by the water level inside the channel (and thus by the channel Manning coefficient) instead of the floodplain Manning coefficient.

In Figs. 5 and 6, we plot the difference between the computed and the observed water levels where image information is available and when considering 1, 3 and 10 land-cover classes. Figure 7 shows the comparison between simulated and observed discharge hydrographs at the EDF middle gauge station (in function of time). These results show that after the identification of an "optimal" set of Manning coefficient values, the computed flow state is much closer to the observations than that computed using the forward hand-calibrated model (i.e. with n values resulting of the trial-error tests-"hand" calibration).

6. Sensitivity Analysis and Manning Decomposition

The previous section shows that our variational data assimilation process is capable of enhancing significantly the calibration of the distributed Manning friction coefficients, and consequently the accuracy of the numerical model. Nevertheless, all computations have been done with an *a priori*: the Manning coefficients have been defined spatially distributed according to given land-use classes (between 1 and 10). This *a priori* is rather traditional. Nevertheless, does this necessary lead to the most accurate numerical model? In the current section, we aim at answering to this question. To meet this aim, a sensitivity analysis is performed without any *a priori* for the Manning spatial distribution (i.e. any land-use is defined).

One run of the forward model plus the adjoint model, without running the full optimization process, gives the gradient of cost function J with

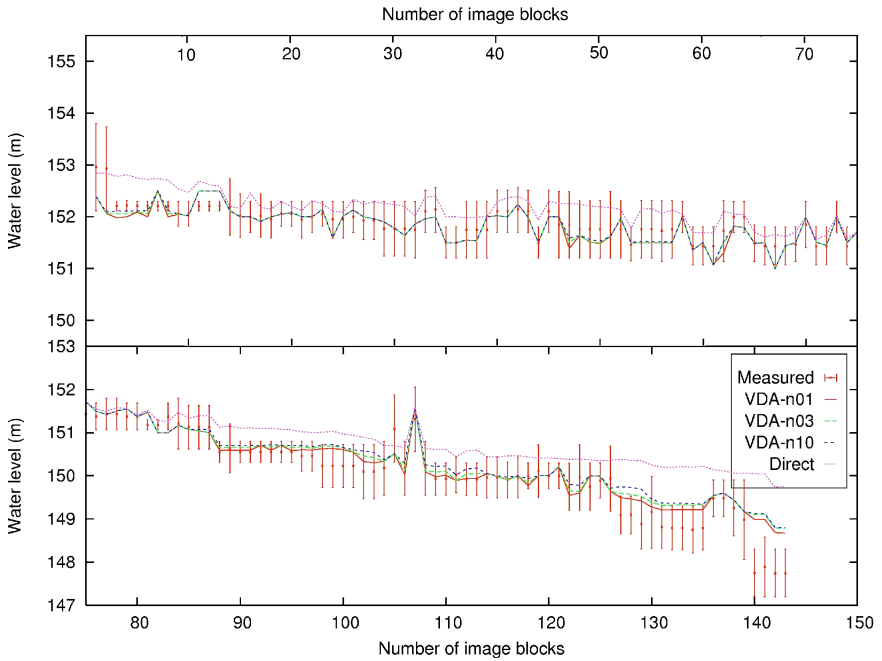

Fig. 5. Water levels (m) at image blocks, i.e. where image information is available (see Fig. 3). Vertical bars correspond to the measures with estimated uncertainties. The four curves correspond to the preliminary forward run and the three calibrated model responses (depending on the land classes number considered; VDA-nXX = 1, 3 and 10 land classes) (Recall: VDA = Variational Data Assimilation = 4D-VAR algorithm; direct = forward model without VDA ("hand" calibrated model).

respect to the Manning coefficients. These gradient values represent the local sensitivity of the cost function with respect to the Manning coefficients. They help one to understand which Manning area is the most important to calibrate.

First, we perform sensitivity analysis in the case the Manning is decomposed by the 10 land-uses defined previously, see Fig. 4. We show the 10 spatially distributed gradient values in Table 1. This sensitivity analysis suggests that the most important Manning value to focus on is in the main channel, than much less important are the vegetation area, the bridge (in main channel), gravel area and grassland area. Others land-use values are negligible.

This sensitivity analysis result is consistent with the calibration process presented in previous section: the optimization algorithm calibrate essentially the main channel value.

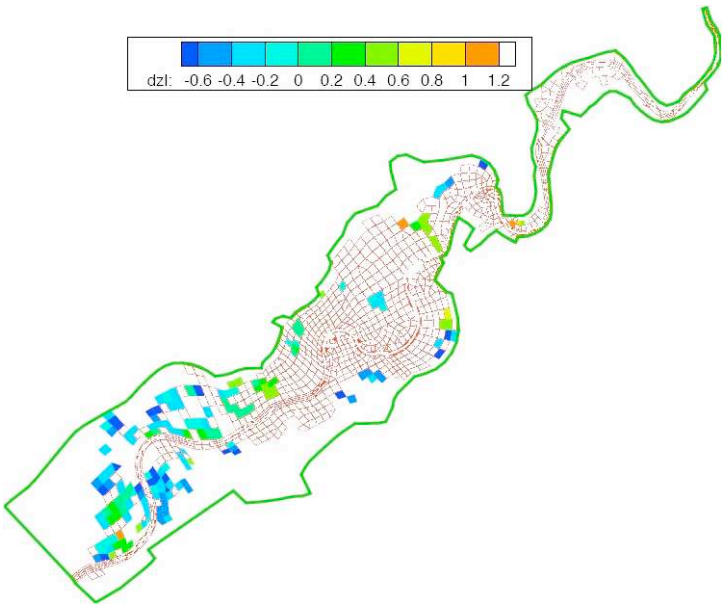

Fig. 6. Difference of water depth (m) at image instant (6:00, Feb 28, 1997) between observations and computed values using the calibrated model (Manning is calibrated in three land classes, Case b).

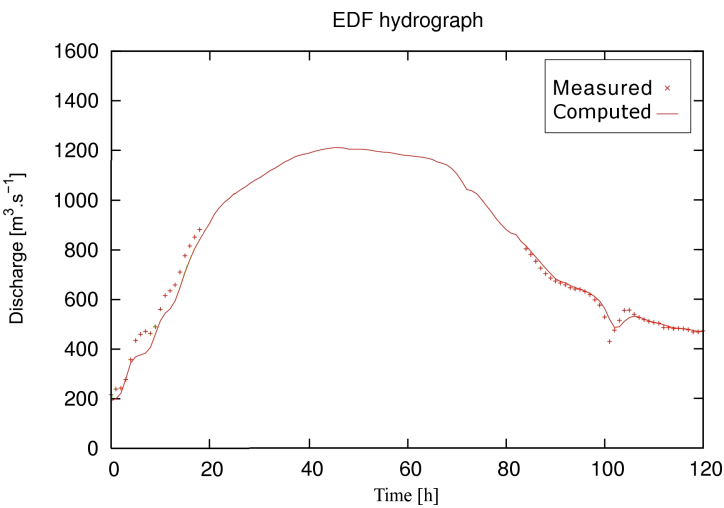

Fig. 7. Measured discharge hydrographs (m^3s^{-1}) at the middle gauge station EDF, and computed one using the calibrated model (Manning is calibrated in three land classes, Case b; at downstream, elevation is imposed).

Table 1. Sensitivity analysis. Gradient values for 10 land-uses (at Manning value constant everywhere $n = 0.033$).

Channel	Vegetation	Bridge	Gravel	Grassland
881.54 e + 6	43.75 e + 6	17.76 e + 6	3.64 e + 6	−1.06 e + 6
Left	Right	Snaked land	Urban	Downstream channel
0.07 e + 6	0	0	0	0

No a priori land-use case

Next, a sensitivity analysis with no *a priori* on the Manning decomposition has been performed: we define one Manning value for each mesh cell, and we compute the gradient value for each cell. In particular in the main channel, there are as many potential values as cells. The computed spatially distributed gradient values are shown in Fig. 8.

The present sensitivity analysis suggests defining few Manning areas inside the main channel (at least four in the present case). More generally,

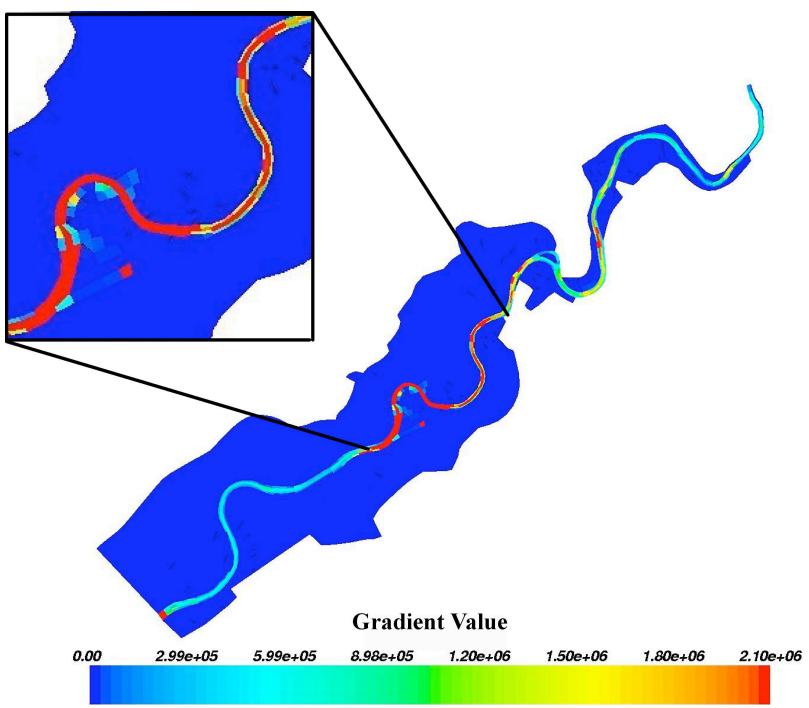

Fig. 8. Sensitivity analysis. Gradient value: cost function with respect to the Manning coefficient in each finite volume cell (there is no *a priori* land-use).

this suggests to define the Manning areas not upon the land-use (or land-cover) but upon "sensitive areas". Furthermore, the combination of the satellite image and runs of sensitivity analysis indicates the most important Manning areas we need to calibrate (including inside the main channel).

7. Conclusion

By using a variational data assimilation approach (4D-var), we have investigated potential contributions of SAR derived spatially distributed water levels for the identification of time-independent parameters (Manning coefficients) in a shallow-water flood model.

The spatially distributed water levels have been derived from a SAR image by employing the method developed in Ref. 12. They have been obtained with a ± 40 cm mean uncertainty, using a RADARSAT-1 image of the 1997 flood event of the Mosel river. This has been possible by using both an analysis of the relevance of SAR derived flood extent limits for hydraulic purposes, and a merge between the relevant limits and a highly resolution DEM, under hydraulic coherence constraints inspired from Refs. 11 and 14. Such a water level estimation provides spatially distributed information at the time of a satellite overpass while classical *in situ* measurements are punctually available.

Numerical experiments conducted in this study indicate that a rather dense information in space is of great benefit for the identification of unknown parameters (Manning friction coefficient). Indeed, the assimilation of the SAR derived water levels, in addition to an incomplete discharge hydrograph, proves to be capable of identifying Manning friction coefficients, while the ground data alone does not allow such an identification. Furthermore, a sensitivity analysis conducted by using the SAR derived water levels, shows that a spatial distribution of the friction coefficient based on land-cover may not necessarily lead to the better model results. Indeed, these water levels, used as a guide in the sensitivity analysis, can define areas of Manning friction homogeneity, without apparent link with land-cover. Such sensitivity analysis may finally base the Manning parameter spatial distribution more on the model hydraulic functioning, than on the land-cover.

In a near future, with the launch of new radar satellites with better spatial and radiometric resolutions and more suitable wavelength, the uncertainties of water levels estimates will presumably be further reduced.

The capability of SAR derived water levels may be enhanced to help the identification of model parameters using variational assimilation.

References

1. G. Schumann, R. Hostache, C. Puech, L. Hoffmann, P. Matgen, F. Pappenberger and L. Pfister, *IEEE Transactions on Geoscience and Remote Sensing* **45** (2007) 1715.
2. Y. Ding, Y. Jia and S. Wang, *Journal of Hydraulic Engineering* **130** (2004) 501.
3. E. Belanger and A. Vincent, *Journal of Hydrology* **300** (2005) 114.
4. H. Roux and D. Dartus, *J. Hydraul. Res.* **43** (2005) 311.
5. W. Castaings, D. Dartus, M. Honnorat, F.-X. Le Dimet, Y. Loukili and J. Monnier, *Lecture Notes in Computational Science and Engineering* **50** (2006) 249.
6. M. Honnorat, J. Marin, J. Monnier and X. Lai, *Research report INRIA RR-6150* (2007).
7. I. Gejadze and J. Monnier, *Comput. Meth. Appl. Mech. Eng.* **196** (2007) 4628.
8. M. Honnorat, J. Monnier and F. Le Dimet, *Comput. Visual. Sci.* **12** (2009) 235.
9. J. Marin and J. Monnier, *Math. Comput. Simul.* **80** (2009) 547.
10. M. Honnorat, J. Monnier, N. Rivière, E. Huot and F. Le Dimet, *Comput. Visual. Sci.* **13** (2010) 111.
11. D. Raclot, *International Journal of Remote Sensing* **27** (June 2006) 2553.
12. R. Hostache, C. Puech, G. Schumann and P. Matgen, *Revue Télédétection* **6** (2006) 325.
13. R. Hostache, X. Lai, J. Monnier and C. Puech, *J. Hydrol.* (2010).
14. C. Puech and D. Raclot, *Hydrological processes* **16** (2002) 1593.
15. X. Lai and J. Monnier, *J. Hydrology* **377** (2009) 1.